DOCUMENT RESUME

ED 397 419 CS 215 352

AUTHOR Callahan, Susan
TITLE Using Portfolios for Accountability: The Ethic of Care Collides with the Need for Judgment.
PUB DATE Mar 96
NOTE 17p.; Paper presented at the Annual Meeting of the Conference on College Composition and Communication (47th, Milwaukee, WI, March 27-30, 1996).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Educational Cooperation; Ethics; Ethnography; High Schools; Language Arts; *Portfolio Assessment; Portfolios (Background Materials); Public Schools; *School Culture; *State Regulation; *State Standards; *Student Evaluation; Teacher Student Relationship; *Writing Evaluation
IDENTIFIERS *Kentucky Education Reform Act 1990

ABSTRACT

In 1989, the Kentucky Supreme Court determined that the state's public school system was unconstitutional and called for the General Assembly to "recreate and re-establish a new system of common schools." A new system of accountability in language arts departments required portfolio assessment that was both objective and strictly consistent with procedures and regulations prescribed by the legislature. An ethnographic study examined this assessment process and explored the ways in which one high school English department, consisting of nine women, found themselves struggling with competing loyalties or ethical commitments. Teachers were not to grade any portfolios containing writing familiar to them, i.e., writing written for them in one of their classes. They were responsible for scoring 23 anonymous portfolios Difficulties began when a third reading was required to resolve a dispute between the first and second ratings. The size of the school made it almost unavoidable that the third reader would have to evaluate a portfolio put together by a former student. Problems in scoring portfolios arose when teachers were required to give low scores to students they knew to be proficient writers and when teachers encountered a portfolio they believed had not been completed by a student. The primary weakness of the Kentucky portfolio assessment system may be that it encourages teachers to practice a dangerous form of self-deception: they are asked to pretend that they can become dispassionate readers, able to evaluate themselves and their students objectively. (Contains 11 references; the Kentucky Writing Assessment holistic scoring guide is appended.) (TB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 397 419

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
S Callahan
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Susan Callahan
Northern Illinois University
CCCC 1996

Using Portfolios for Accountability: The Ethic of Care Collides with the Need for Judgment

In 1989, the Kentucky Supreme Court determined that the state's public school system was unconstitutional and called for the General Assembly to "'re-create and re-establish a new system of common schools'" (Legislative Research Commission, 1991). In response, the General Assembly passed the Kentucky Education Reform Act, designed to correct inequities in finance, governance, and curriculum. This act, known throughout the state as KERA, is one of the most dramatic and ambitious reform efforts currently under way in the country, and its impact is being closely observed by other states hoping to learn from Kentucky's efforts. To date, although the scope of the finance and governance provisions of KERA have rocked many Kentucky school districts, the greatest impact has come from the curriculum reform portion of the bill since it has affected every student, teacher, and administrator in the state.

Changing curriculum is not easy, and the General Assembly, in recognition of this fact, mandated the creation of a new accountability system that could be used to stimulate and measure the desired curriculum changes. Consequently, KERA required the State Board of Education to develop a statewide assessment program, using performance tests in place of the standardized tests then being used. One of these new tests was to be writing portfolios collected from all fourth, eighth, and twelfth grade students. To ensure that teachers and administrators took these assessments seriously, KERA specified that test scores would be used to determine financial rewards for teachers in successful schools and sanctions against schools that were

CS 215 353

BEST COPY AVAILABLE

not successful (Foster, 1991).

In using performance assessment to drive school reform, the Kentucky State Board of Education hoped it had discovered what Grant Wiggins (1992) calls "tests worth taking"--that is, tests that require the kinds of complex higher order thinking and problem solving skills not measured by standardized tests and that require students to demonstrate they can use their knowledge in contexts closely mirroring real life situations. When the General Assembly voted to use performance tests for accountability, however, they anticipated the tests would do more than just provide significant information about Kentucky students. They intended the tests to improve teaching as well. In fact, they hoped that the teachers would teach to the tests so that the tests would shape and improve both curriculum and pedagogy (Foster, 1991). The Department of Education was particularly outspoken about the expectation that the new tests would "be a way of changing curriculum to encourage more writing and process guided instruction" (T. Wilson, personal interview, April 21, 1992).

Since KERA was funded by a hefty tax increase, the Kentucky General Assembly wanted the initial set of test scores as quickly as possible so that they would have a baseline record against which to measure future growth. Consequently, the Department of Education had only a little over a year in which to design, explain, and implement the new assessment system. To say that the first year of the assessment was chaotic would be an understatement. Near the end of the 1991-1992 school year, Tish Wilson, the woman who was then the Writing Program Coordinator at the Department of Education, described the experience as "building the boat while we sail it," and that image comes fairly close to capturing the frantic energy, confusion, and anxiety that engulfed the Department of Education and flowed downward throughout the entire

school system.

The tension surrounding the new tests was exacerbated by the fact that good test scores hinged upon writing, a skill that long had been under-valued and under-taught in Kentucky. Students who were used to classes that required little or no writing were now required to assemble portfolios that included six compositions, two of which had to come from classes other than English; and they were also required to use writing to explain their answers on the new performance tests in science, math, reading, and social studies. As a result, writing instruction suddenly achieved a prominence in Kentucky that it had never had before.

As a writing teacher, I naturally wanted to see how teachers were understanding the new portfolio requirement and interpreting it to their students, so I designed an ethnographic study of one high school English Department and observed and participated in their school, which I will call Pine View, throughout the second year of the portfolio assessment. The department consisted of nine women, most of whom had worked together for between eight and thirteen years. Eight had Master's Degrees in English Education, and three had received some additional experience with the Blue Grass Writing Project. Although the women differed from each other in many ways, they described themselves as an "unusually close" department, and they were. The main reason they worked so well together was that they all agreed their job was to do "whatever is best for the kids" and often saw themselves as the sole line of defense between their students and some form of bureaucratic injustice. Departmental leadership was collaborative, and decisions were based on doing whatever they felt would help their students succeed individually and collectively. In short, this all female department operated within a strong feminine ethic of care for their students and for

each other (Gilligan, 1982).

The speed with which the portfolio assessment was developed and implemented meant that the Department of Education was not able to provide teachers and administrators with the time and information they needed to understand the assessment before it was put into place. During the first two years following KERA, information was distributed primarily from the Department of Education, to district coordinators, to "cluster leaders" (teachers who were to explain the portfolio content requirements and scoring system to teachers in their school or to teachers in a cluster of schools). Consequently, the Pine View English teachers' initial understanding of the portfolio assessment was in large part determined by the woman who was their cluster leader. This teacher, Lynn, attended meetings and workshops and passed along whatever information and materials she received, all of which stressed the importance of the evolving set of guidelines and regulations for portfolio scoring.

Lynn was an experienced and dedicated teacher who saw the new portfolio requirement as a professional challenge. She was proud of her department's well deserved good reputation and was determined to maintain it by correctly interpreting and enforcing all the rules for portfolio assembly and scoring. In this, she was fully supported by the other members of the department who trusted her judgment and admired her intelligence and high standards. The teachers were determined to score the portfolios accurately, believing that this was the best way to demonstrate both their department's high standards and the need for students to receive more frequent and better writing opportunities in their non-English classes.

During the course of my study, I learned a great deal about the dynamics of the Pine View English Department and about the wider Pine View school

culture; I also saw how these various elements affected the way the portfolio concept was understood and implemented within this particular school. I observed some of the ways concepts and guidelines were distorted as they were passed down from the portfolio assessment designers through several layers of administrators and teachers to the high school seniors who were required to compile portfolios in order to graduate, and I noted how the legislative emphasis on portfolio scores prevented the teachers from realizing many of the pedagogical benefits the Department of Education intended to encourage. Since these findings have been discussed elsewhere (Callahan, in press), I will illustrate the problems inherent in asking teachers to help assemble and score the student portfolios that will be used as the teachers' accountability instrument by focusing on one significant difficulty the Pine View teachers had during portfolio scoring. This difficulty arose out of the state's expectation that the teachers would be able to evaluate portfolios compiled by Pine View seniors without allowing their familiarity with those students or their own positions as faculty members within Pine View to become a part of the evaluation process.

When the English teachers met in March to score the 207 Pine View portfolios, Lynn reviewed all the regulations and suggestions she had received from the Department of Education, but she stressed two principles above all others: (1) that portfolios must be scored by adhering strictly to the language of the scoring guide so that the scorer and/or the entire department would not be found "discrepant" and reported to the administration, and (2) that scorers were supposed to score only portfolios containing work by students they did not know. Both of these principles became harder and harder to follow as the portfolio scoring progressed.

During the first day of scoring, Lynn made certain that each teacher

was given responsibility for scoring 23 anonymous portfolios. Each portfolio was identified only by a number, and if a teacher discovered she was reading familiar material, she was instructed to exchange that portfolio for an unfamiliar one. On the second day of scoring, the teachers each read a previously scored collection of portfolios with no knowledge of how these portfolios had been rated. Once Lynn had compared the two scores, everyone was surprised and pleased to discover that carefully following the scoring guide for each of the four possible categories had resulted in agreement on most of the portfolios (For a copy of the scoring guide, see Appendix A). The remaining pile of thirty-three portfolios, those that Lynn called the "thorny" ones, required a third reading. Since Lynn had received the message that teachers "should" be able to reach consensus on portfolio scores, she saw this expectation as a professional challenge and insisted that the teachers attempt to reach agreement on all of the thorny portfolios by discussing them in the language of the scoring guide. This kind of detached analysis became increasingly difficult, however, once the teachers were serving as the second or third reader of a portfolio, largely because second and third readers frequently knew the student whose work was being discussed. Such knowledge was inevitable since five of the nine teachers who were scoring the portfolios taught at least one section of senior English.

Of course, even during the first reading day teachers occasionally came across a familiar piece of writing within a collection of unfamiliar work. Usually these pieces presented no real problem because the familiar piece was consistent with the other work in the portfolio, and frequently the teacher remembered the writing from a past semester but not the student who had written it. When the portfolios were being read and discussed for the second or third time, however, such discoveries became more frequent

and more disturbing.

Some tension between the mandate to score objectively in the language of the scoring guide and the contextual knowledge they had of Pine View students and faculty arose when teachers were forced to give low scores to portfolios compiled by students they knew were proficient writers but who had misunderstood directions, included weak pieces based on poorly conceived assignments from classes outside the English Department, had not edited some of their pieces, or had simply chosen to ignore certain requirements. When these problems resulted in Incomplete portfolios, the teachers extended the portfolio deadline to pressure students to make changes and scored the re-submitted portfolios after the official scoring days were over. Even though this effort increased the time and energy they needed to spend on the assessment, they believed this effort was necessary because each Incomplete portfolio would lower the overall Pine View accountability score. They did, however, reluctantly give low scores to complete but flawed portfolios, including those written by students they knew were capable of better work but whose portfolios, for whatever reason, did not demonstrate this ability. They also gave low scores to portfolios from students who did indeed write poorly, but often these low scores were accompanied by verbal complaints that the score did not reflect the physical, emotional, and intellectual constraints under which that individual student was working.

But most of the tension arose when teachers encountered a portfolio they believed had not been honestly compiled by a student. Although teachers had been warned to look for plagiarism and to score those portfolios as Incomplete, the Department of Education's requirement that they must be able to document the original source made it difficult for the teachers to substantiate their belief that some students had borrowed work from friends

or had received considerable parental help in polishing portfolio pieces. While these practices did not seem to be widespread, both teachers and students acknowledged that they had occurred. The teachers did disqualify a few portfolios that contained obvious re-tellings of familiar movie or television plots, but portfolios containing pieces that were questionable only because the teachers were familiar with the student writers were not disqualified when their problematic nature would not be apparent to an outside reader. These portfolios could be, and were, scored only in terms of the language of the scoring guide.

I remember one portfolio in particular that generated a great deal of discussion before it was given its final score. Most portfolios that required a third reading were ones where readers disagreed whether a portfolio was Novice or Apprentice, the two lowest categories, or where they disagreed between the mid-range scores of Apprentice and Proficient. The few Distinguished portfolios were usually easy to identify because, as one teacher put it, "They jump out at you." This portfolio, though, was rated as Distinguished by its first reader and high Proficient by its second, who felt that the language was stilted and that some of the vocabulary words seemed to have been selected by browsing through a thesaurus. The portfolio went to Lynn for its third reading. This presented a difficulty, though, because Lynn knew the student who, in fact, was in her AP English class. She asked another teacher to read the portfolio, but that teacher, too, was working with the student in an elective English course. Finally, they both agreed to read the portfolio because everyone else was busy, hoping their scores would be in agreement with either the first or second reader.

But when Lynn began reading the portfolio, she recognized that several of the pieces had undergone significant stylistic changes since she had seen

them in class. Evidently surprised, she said, "Oh, no! Her mother is a teacher, and I'll bet she helped her with this." The second teacher responded, "Yes, she is always pressuring her." The conversation immediately shifted away from the portfolio to the student and her relationship with her mother. A couple of other teachers joined in because they knew the student and/or the mother. They all were more concerned about what would happen to the student if her portfolio was not rated as Distinguished than with whether or not it should be disqualified. Eventually the portfolio was read and discussed by most of the teachers, but each reading made the scoring decision more difficult. Finally, the original scorer, who did not know the student, was told to use her own judgment of how the portfolio would be perceived by someone from the Department of Education. She retained her original scoring decision, and the portfolio was designated as Distinguished.

This incident illustrates just one of the ethical gray areas that the teachers experienced during the entire portfolio process, where one set of loyalties was pitted against other equally compelling loyalties, and where one standard of excellence was in conflict with another. Much of the discomfort the teachers experienced during portfolio scoring was due to their determination to do the "right" thing. By this, I do not mean just their desire to label each portfolio correctly according to the standards of the scoring guide--although the pursuit of that putative "right" score was certainly the major emphasis of their reading. Neither do I mean just their desire to correctly interpret every guideline and requirement leading up to the two days of portfolio reading although this, too, had been a significant concern. Rather, I mean their desire to do the right thing for their students, their school as a whole, and for themselves as a department--all in relation to their support for the Department of Education's

intent to use the portfolios to improve education for all Kentucky students.

Making such a singular "right" decision was, of course, impossible. Instead, they were forced to make a series of compromise decisions, sometimes letting one set of concerns become primary, and then later balancing that decision by focusing on another set of needs. Making such compromises, however, meant they felt uncomfortable with most of the decisions they made. Consequently, they kept turning away from the hierarchical decision making system set forth by the Department of Education to incorporate the reality of the students and teachers whose lives would be affected by their decisions. Then, they would turn back to the language of the scoring guide for relief from the pressure of having too much information that they were unable to communicate within the assessment format.

Throughout the scoring procedure, they not only experienced what Brian Huot (1990) calls the tension between "reader as reader and reader as rater (p. 225) stemming from their experience as English teachers (Barritt, Stock, & Clark, 1986), but they experienced the additional pressure of trying to read as if they had no knowledge of the specific students who had created the portfolios or the way the resulting portfolio scores would be used to judge their own competence as professionals. As a result, they increasingly used the language of the scoring guide to justify decisions that sometimes were rooted in factors that had little to do with the observable quality of the work in the portfolio being rated. At the same time, they realized that some of their most apparently "objective" decisions did not reflect their best professional judgments.

In the past, externally mandated tests had been experienced as something imposed upon both teachers and students, who had stood together outside a testing event that was unrelated to their regular relationship. The portfolio

assessment, however, not only required the teachers to enthusiastically promote the lengthy portfolio assembly process as a worthwhile activity, it also required them to evaluate their students and themselves according to the criteria established by an outside authority for purposes established by that authority.

By using portfolios for accountability, the state of Kentucky has upset many of the delicate balances that used to exist between pedagogy and evaluation and between internal and external assessment. In Embracing Contraries, Peter Elbow (1986) suggests that the only way a classroom teacher can work well is to keep separate her two opposing functions--the one that requires her to serve as guide and supporter and the other that requires her to be an advocate of knowledge and standards. She can do this by never attempting to fulfill these contrary functions at the same time. Unfortunately, the accountability system developed by the state of Kentucky has conflated these two roles. The senior English teachers, in particular, know that it is to their advantage to do whatever they can in their classrooms to see that students submit portfolios that can be evaluated favorably in the language of the scoring guide. Since they are encouraged to design writing assignments with the assessment criteria in mind and to discuss writing in the terms used in the scoring guide, assessment is, indeed, beginning to change the English curriculum. For this reason, however, the portfolio system poses ethical problems far beyond the considerations Elizabeth Flynn (1989), Nel Noddings (1984), Betty Shiffman (1992), and others have described as inherent in the practice of assigning grades to student writing.

Although the portfolio assessment has, indeed, stimulated a greater amount of student writing, it has made it more difficult for teachers to

separate student purposes for writing from teacher purposes for having students write. Additionally, it has encouraged teachers to see student texts as products that can and should be viewed objectively. Moreover, teachers are not able to counteract this objectification of student texts--and students--by turning the task of evaluation over to external examiners because the teachers themselves have become the external examiners. Ultimately, then, the primary weakness of the Kentucky portfolio assessment system may be that it encourages teachers to practice a dangerous form of self deception: They are asked to pretend that they can become dispassionate readers, able to evaluate themselves and their students objectively.

References

Barritt, L., Stock, P., & Clark, F. (1986). Researching practice: Evaluating assessment essays. College Composition and Communication, 37, 315-327.

Callahan, S. (in press). Kentucky's state-mandated writing portfolios and teacher accountability. In I. Weiser & K. Yancey (Eds.), Situating Portfolios: Four perspectives. Utah State University Press.

Elbow, P. (1986). Embracing contraries. New York, Oxford University Press.

Flynn, E. A. (1989). Learning to read student papers from a feminine perspective, I. In B. Lawson, Susan Sterr Ryan, & W. Ross Winterowd (Eds.), Encountering Student Texts: Interpretive Issues in Reading Student Writing (pp. 49-57). Urbana, IL: NCTE.

Foster, J. D. (1991). The role of accountability in Kentucky's education reform act of 1990. Educational Leadership, 48(5), 34-36.

Gilligan, C. (1982). In a different voice: Psychological theory and womens's development. Cambridge, MA: Harvard University Press.

Huot, B. A. (1990). The literature of direct writing assessment: Major concerns and prevailing trends. Review of Educational Research, 60, 237-263.

Legislative Research Commission. (1991). The Kentucky education reform act of 1990: A citizen's handbook. Frankfort, Kentucky.

Noddings, N. (1984). Caring: A feminine approach to ethics and moral education. Berkeley, CA: University of California Press.

Shiffman, B. G. (1992). Reading our students/ourselves: Toward a feminist theory of evaluation. Teaching English in the Two Year College, 19, 24-31.

Wiggins, G, (1992). Creating tests worth taking. Educational Leadership,

49(8), 26-33.

KENTUCKY WRITING ASSESSMENT
Holistic Scoring Guide

1992-93

NOVICE	APPRENTICE	PROFICIENT	DISTINGUISHED
• Limited awareness of audience and/or purpose • Minimal idea development; limited and/or unrelated details • Random and/or weak organization • Incorrect and/or ineffective sentence structure • Incorrect and/or ineffective wording • Errors in surface features are disproportionate to length and complexity	• An attempt to establish and maintain purpose and communicate with the audience • Unelaborated idea development; unelaborated and/or repetitious details • Lapses in focus and/or coherence • Simplistic and/or awkward sentence construction • Simplistic and/or imprecise language • Some errors in surface features that do not interfere with communication	• Focused on a purpose; evidence of voice and/or suitable tone • Depth of idea development supported by elaborated, relevant details • Logical organization • Controlled and varied sentence structure • Acceptable, effective language • Few errors in surface features relative to length and complexity	• Establishes and maintains clear focus; evidence of distinctive voice and/or appropriate tone • Depth and complexity of ideas supported by rich, engaging, and/or pertinent details; evidence of analysis, reflection, insight • Careful and/or subtle organization • Variety in sentence structure and length enhances effect • Precise and/or rich language • Control of surface features

Analytic Annotation Guide

CRITERIA	OVERVIEW	COMMENDATIONS		NEEDS	
PURPOSE/ APPROACH	The degree to which the writer • establishes and maintains a purpose • communicates with the audience	P/A-X	clear awareness of audience and purpose	P/A-J	greater sense of audience and purpose
		P/A-Y	original and/or insightful approach and evidence of distinctive voice/tone	P/A-K	more insightful approach and evidence of voice/tone
IDEA DEVELOPMENT/ SUPPORT	The degree to which the writer provides thoughtful, detailed support to develop the main idea(s)	I/S-X	perceptive thinking	I/S-J	more thoughtful investment by author
		I/S-Y	relevant, interesting details	I/S-K	more elaboration of details
ORGANIZATION	The degree to which the writer demonstrates • logical sequencing • coherence • transitions/organizational signals	OX	evidence of planning	OJ	more evidence of planning
		OY	order/sequence easily followed	OK	more logical sequence of ideas and effective transitions
SENTENCES	The degree to which the writer includes sentences that are • varied in structure and length • constructed effectively • complete and correct	SX	variety in structure and length	SJ	greater variety in structure and length
		SY	effectively constructed sentences	SK	more effective sentence construction
WORDING	The degree to which the writer exhibits correct and effective • word choice • usage	WX	successful use of pertinent and/or rich language	WJ	closer attention to effective word choice
		WY	control of conventional usage	WK	greater control of conventional usage
SURFACE FEATURES	The degree to which the writer demonstrates correct • spelling • punctuation • capitalization	SF-X	spelling enhances readability	SF-J	accurate spelling
		SF-Y	capitalization and punctuation aid clarity	SF-K	greater control of punctuation and capitalization

BEST COPY AVAILABLE